Appendix for

Bride Price and Female Education

Nava Ashraf

London School of Economics and *BREAD*

Natalie Bau

University of California, Los Angeles, *CEPR*, and *CIFAR*

Nathan Nunn

Harvard University, *NBER*, and *BREAD*

Alessandra Voena

University of Chicago, *NBER*, *CEPR*, and *BREAD*

August 20, 2018

Contents

Appendix A: Data description

Indonesian Data Sources

1995 Intercensal Survey

The Indonesia Intercensal Survey is a large-scale, nationally representative population survey of Indonesia carried out between the 1990 and 2000 censuses. It is housed by the Minnesota Population Center (1995). Importantly, it includes data on primary language spoken, which can be linked to ethnicity and matched to an ethnic group's bride price custom in the *Ethnographic Atlas*. It also contains information on educational attainment, birth year, and birth district, which following Duflo (2001), can be combined with data on the number of schools built in 1974 as part of the large-scale school construction program.

Indonesian Family Life Survey (IFLS)

The Indonesian Family Life Survey (IFLS) is an ongoing longitudinal study of households in Indonesia covering over 30,000 individuals. Data is gathered from 13 of Indonesia's 27 provinces, and the study is considered representative of 83 percent of the Indonesian population. This paper uses data from rounds 3 and 4 of the IFLS (Strauss et al., 2004, 2009), which unlike previous rounds of the IFLS, includes questions about individuals' ethnicities.

2010 Census

We use a 10 percent random extract of the 2010 Indonesia Census to replicate the results on the effects of the school construction program in a second, larger data set. This extract is also housed by the Minnesota Population Center. We use a concordance between the ethnic groups in the census and the *Ethnographic Atlas* to determine an individual's ethnicity-level customs. Our sample of interest consists of the same cohorts analyzed in the Intercensal data (a treated group of individuals who were 2–6 in 1974, at the time of school construction, and an un-treated group of individuals who were 12–24 at the time of school construction).

Zambian Data Sources

Data from the Zambia Fertility Preferences Study (ZFPS)

Data on bride price amounts and beliefs about bride price and education are drawn from unique survey data collected in Lusaka in Fall 2014 as part of an experimental study on family planning. The study involves 715 couples (1,430 individuals) living in the catchment area of Chipata clinic, a poor, peri-urban segment of Lusaka. Each spouse was interviewed in private and was asked a series of questions on the practice of *lobola*, leading to a total of 1,430 observations.

Demographic and Health Survey (DHS)

The Zambian *Demographic and Health Survey* is a nationally representative survey of men and women in reproductive age, and includes data on respondents' children. Our analysis pools the 1996, 2001, 2007, and 2013 rounds of the survey. The DHS contains data on school-aged children's school enrollment and ethnicity, which can be matched with data on bride price traditions from the *Ethnographic Atlas* and the stock of schools from the Zambian Ministry of Education. The pooled surveys also contain data on gender attitudes from both male and female adult respondents, allowing us to test whether bride price is correlated with gender attitudes.[1] To do so, we form two indices, pooling male and female respondents. The first index is the portion of times a respondent replied that a husband was *not* justified in beating his wife, and the second index is the portion of times a respondent said that a wife was justified in refusing a husband sex.[2]

Ethnographic Data Sources

Information on bride price practices is taken from the *Ethnographic Atlas* (Murdock, 1967) and LeBar (1972) for Indonesia and the *Ethnographic Atlas* (Murdock, 1967) and the *Ethnographic Survey of Africa* (Willis, 1966; Whiteley and Slaski, 1950; Schapera, 1938) for Zambia.

Our primary analysis for Indonesia uses the 1995 Indonesia Intercensal Survey, which records 174 different spoken languages. These are matched to 44 ethnic groups from the *Ethnographic Atlas*

[1] The gender attitude questions were not asked in the 1996 DHS.

[2] The questions for the first index take the form, "Wife beating justified if..." The options were "if she goes out without telling him," "if she neglects the children," "if she argues with him," "if she refuses to have sex with him," and "if she burns the food." The questions in the second index take the form, "reason for not having sex:", and the possible answers are "husband has STD," "husband has other women," "recent birth" (only asked in 2001) and "tired, mood."

and LeBar (1972). To undertake the matching, we exploited the previous matching of ethnic groups to languages undertaken by Alesina et al. (2013), where the 1,265 ethnic groups of the *Ethnographic Atlas* were matched to one of 7,612 language groups in the *Ethnologue: Languages of the World* (Gordon, 2005). All but 11 of the 172 language groups in the Indonesia Intercensal Survey could be matched to an ethnicity from our sources. These comprise 0.43 percent of the observations with non-missing language data.

Our baseline analysis for Zambia uses the four rounds of the Zambia Demographic and Health Surveys (1996, 2001, 2007, and 2013). The Zambia DHS reports 65 distinct ethnic groups. Of these, we are able to match 53 of them to 30 more-coarsely defined ethnic groups from the *Ethnographic Atlas* and the *Ethnographic Survey of Africa* (Willis, 1966; Whiteley and Slaski, 1950; Schapera, 1938). The remaining unmatched groups are small and comprise less than 2.5 percent of the DHS sample. The matching was done by hand, relying heavily on Murdock (1959).

The ethnicity-level control variables used in the analysis (female participation in agriculture and matrilineal descent) are taken from the *Ethnographic Atlas*. We create a measure of female participation in agriculture using variable $v54$ ("sex differences: agriculture") from the *Ethnographic Atlas*. The original variable records ethnic groups as belonging to one of the following categories: (1) males only, (2) males appreciably more, (3) differentiated but equal participation, (4) equal participation, (5) female appreciably more, (6) females only, and (7) absent or unimportant activity. Using this information, we create a measure of female participation in agriculture that takes on the value of one for categories 5 and 6 and zero for all other categories. We also create a second indicator variable that equals one if either agriculture was not present or was unimportant – i.e., category (7) – or if information was missing for the ethnicity. This variable captures the fact that, in both cases, we are not able to observe female participation in agriculture.

A second control variable is a measure of the presence of matrilineal inheritance. This was based on variable $v43$ ("descent: major type"). The original variable groups ethnicities into one of the following categories: (1) patrilineal, (2) duolateral, (3) matrilineal, (4) quasi-lineages, (5) ambilineal, (6) bilateral, and (7) mixed. We construct a matrilineal indicator variable that takes on the value of one if variable $v43$ identifies an ethnic group as belonging to category 3, and zero otherwise. Information on inheritance is never missing if bride price information is available.

Appendix B: Additional analyses

Auxiliary prediction: Bride price and the ability of educated girls

The model delivers an additional testable prediction about the average ability of girls who are educated for bride price and non-bride price groups. Suppose that the support of the ability distribution is $[\underline{a}, \bar{a}]$. The average ability of educated girls is equal to

$$E[a_i|S = 1, I_e, k] = E[a_i|a_i > a^*_{I_e}(k)] = \int_{a^*_{I_e}(k)}^{\bar{a}} a_i g(a_i|a_i > a^*_{I_e}(k)) da_i.$$

By the Leibniz integral rule, $\frac{\partial E[a_i|a_i>a^*]}{\partial a^*} = \frac{g(a^*)}{1-G(a^*)} \{E[a_i|a_i > a^*] - a^*\} > 0$. This gives the following prediction.

Prediction *The average ability of educated girls is lower among ethnicities that engage in bride price payments relative to ethnicities that do not.*

This prediction stems from the fact the fact that enrollment rates are higher among ethnicities that practice bride price (Prediction 3). This, in turn, implies that girls with relatively lower ability are educated in bride price ethnicities.

To test this prediction, we use self-reported, retrospective test score data, taken from rounds 3 and 4 of the IFLS, to test whether this is the case for primary school students in Indonesia. Unfortunately, no similar data are available for Zambia, and so we only have estimates for Indonesia.

Our sample includes all female students in the IFLS who took the Ebtanas state exam, which was the national exam between 1980 and 2001.[3] Our estimating equation is:

$$TestScore_{iepst} = \alpha_t + \delta_{ps} + \beta_1 I_e^{BridePrice} + \mathbf{X}_i\mathbf{\Gamma} + \mathbf{X}_e\mathbf{\Pi} + \varepsilon_{iepst}, \tag{A1}$$

where i indexes a female former primary school student, e indexes her ethnicity, p indexes a province, s indexes the year the test was taken, and t indexes the IFLS survey year. The dependent variable is the student's test score, normalized to have mean of zero and a standard deviation of one in our sample. α_t denotes survey-year fixed effects and δ_{ps} denotes province by test-year fixed effects, which are included because the Ebtanas exam system was standardized at the province-level each year. The vector $\mathbf{X}_i$ includes our age controls, age and age squared, and a control for whether

[3] In 2001, the Ebtanas exam was replaced by the UNAS exam.

the individual reports being Muslim.[4] The vector $\mathbf{X}_e$ includes our set of ethnicity-level controls: indicators for belonging to a traditionally matrilineal ethnic group or an ethnic group with female participation in agriculture. We report standard errors clustered at the ethnicity level, as well as p-values from a wild cluster bootstrap procedure.

Estimates of equation (A1) are reported in columns 3 and 4 of Table 5. Column 3 reports estimates without the Muslim and ethnicity controls, while column 4 reports estimates with the controls. Consistent with the auxiliary prediction, the test scores of girls from bride price groups are estimated to be 0.126 standard deviations lower than girls from non-bride price groups.

Instrumental variables regression for the bride price education premium

This sections reports details of the specifications behind the estimates of the impact of education on bride price amounts, using the same procedure as in Duflo (2001), where school construction is used as an instrument for the educational attainment of women from bride price ethnic groups. We restrict our sample to ethnic groups that practice bride price. As we have shown, for non-bride price ethnic groups, there was no impact on female education and therefore no first-stage predictive power. In line with Duflo (2001), we allow the effect of school construction to vary by a child's age in 1974, restricting the effect to be 0 if a child was older than 12 in 1974. Following Duflo (2001), we also restrict the sample to those born between 1950 and 1972. Unfortunately, the resulting sample of couples from bride price ethnicities who were asked questions about bride price is only 264.

The first-stage estimating equation is:

$$I(Completed\ Primary)_{idkt} = \alpha_d + \alpha_k + \alpha_t + \sum_{a=2}^{12} \beta_a\, Intensity_d \times I(age_{1974} = a)_k + \sum_j \mathbf{X'}_d \mathbf{I}_k^j \mathbf{\Gamma}_j + \epsilon_{idkt}, \tag{A2}$$

where d denotes the district, i denotes the individual, t denotes the survey year, and k denotes the cohort. α_d denotes district fixed effects, α_k cohort fixed effects, and α_t survey-round fixed effects. $\sum_j \mathbf{X'}_d \mathbf{I}_k^j \mathbf{\Gamma}_j$ are the cohort-specific controls for the INPRES sanitation program, the enrollment rate in 1971, and the total number of school-aged children in 1971. In addition, we include a quadratic

[4] Since test takers are asked about their scores retrospectively as adults, we typically do not observe test-takers in the same household as their parents. Therefore, we cannot include controls for other household characteristics such as socioeconomic status or whether parents practiced polygyny.

in age of marriage to account for the fact that bride prices are reported in nominal terms.

The second-stage equation is:

$$y_{idkt} = \alpha_d + \alpha_k + \alpha_t + \gamma I(CompletedPrimary)_i + \sum_j \mathbf{X'}_d \mathbf{I}_k^j \mathbf{\Gamma}_j + \mu_{idkt}, \quad \text{(A3)}$$

where the outcome variable y_{idkt} is the natural log of the bride price.

Appendix Table A6 reports the first stage for the 2SLS estimates. The instruments are jointly significant, but are weak (F-statistic of 3.04). In the main text, Table 4 reports the second stage estimate of the effect of primary schooling on log bride price. The point estimate is imprecise, but it corroborates the results from the OLS regressions. Completing elementary school increases bride price payments by a statistically significant 233 percent.

Alternative measures of bride price

We now describe alternative measures of an ethnic group's practice of bride price. Our alternative measures use data on the value of bride price transfers made in marriages that are reported in the IFLS for Indonesia and ZFPS for Zambia. These measures differ from our baseline strategy since they use contemporaneous data rather than a pre-determined measure of traditions. However, we also report estimates where we use the traditional practice as an instrument for the contemporaneous measure. As we will show, all the specifications result in estimates that are qualitatively similar to our baseline estimates.

We begin by estimating the following equation:

$$BP\,Amount_{iet} = \alpha_t + \sum_{j \in E} \alpha_j I_e^j + \beta_1 I_{iet}^{Primary} + \beta_2 I_{iet}^{JuniorSec} + \beta_3 I_{iet}^{SeniorSec} + \mathbf{X_{iet}}\mathbf{\Gamma} + \epsilon_{iet}, \quad \text{(A4)}$$

where i indexes a marriage, e indexes the ethnicity of the wife, and t indexes the survey year. $BP\,Amount_{iet}$ is the monetary value of the bride price, measured in nominal domestic currency. I_e^j is an indicator variable that equals one when ethnicity e belongs to ethnic group $j \in E$. $I_{iet}^{Primary}$, $I_{iet}^{JuniorSec}$, and $I_{iet}^{SeniorSec}$ denote three school completion indicators and are defined as in equation (4). $\mathbf{X_{iet}}$ consists of a quadratic in the bride's marriage age and a quadratic in the bride's marriage year, which helps to account for inflation.

Our coefficients of interest are the α_j's, which provide a measure of the average difference of the value of bride price paid to women belonging to each ethnic group after accounting for the education, marriage age, and marriage year of brides. The estimated α_j's can be interpreted as measures of bride price "value-added" that are analogous to measures of teacher value-added in the education literature (Chetty et al., 2014). For ease of interpretation, we normalize these fixed effects to range from zero to one. This provides a continuous measure of the extent to which each ethnic group practices bride price. We denote this measure $BP\,Continuous_e$.

The estimating equation (A6) has a very similar structure to equation (4) in the paper. However, an important difference is that in equation (4), we were interested in the relationship between the value of the bride price received and schooling when there was a bride price. Therefore, this specification excluded observations for which the value of the bride price was zero and took the natural log of bride price. In this specification, we are interested in classifying all groups, including those who do not practice bride price. Thus, we include observations with a zero bride price and do not take the natural log of the measure.

We then manually match the ethnic groups reported in the IFLS and the ZFPS to the (larger but not necessarily overlapping) ethnic groups in the Indonesian Intercensus and the Zambia DHS, respectively. This allows us to merge our continuous bride price value-added measures to the data that are used to estimate the heterogeneous effects of the school construction program.

There are a number of reasons to be cautious when using the estimate bride price measures. First, modern payments are likely endogeneous to current policies and ethnicity-level characteristics and much more so than traditional practices. Second, the smaller ZFPS and IFLS often contain only a few observations per ethnic group (as few as 8 observations in the IFLS and 2 in the ZFPS).[5] These small sample sizes suggest that measurement error may lead to attenuation bias when the estimated bride price measures are used as explanatory variables.

We undertake two strategies to help address these potential issues. The first is to use the estimates to construct two sets of groups that mimic our baseline bride price indicators: those that practice bride price and those that do not. The alternative bride price indicator variable, $I_e^{AltBridePrice}$, is equal to one if an individual belongs to an ethnic group with an estimated value of α_j that is equal to or above the median value within the population and zero otherwise. Our

[5]When multiple equally-valid matches for an ethnic group in the IFLS or ZFPS were available, we matched to the ethnic group with the greater number of observations in the data. In this way, our constructed bride price measures are based on the largest samples possible.

second strategy is to use our baseline traditional bride price indicator as an instrument for our contemporary measures of bride price.

The estimates are reported in Appendix Table A13. Columns 1–4 report estimates for Indonesia and columns 5–8 report estimates for Zambia. In the odd numbered columns, we report estimates of equations (7) and (10) of the text, but using either $I_e^{AltBridePrice}$ and $I_e^{AltNoBridePrice}$ or $BP\,Continuous_e$ as our ethnicity-level measures of the practice of bride price. We also alter the set of controls and include fixed effects for the ethnic groups in the IFLS or ZFPS. While the odd numbered columns estimate equations (7) and (10) using OLS, the even numbered columns report IV estimates that use our baseline traditional bride price indicator as an instrument.

The estimates reported in Appendix Table A13 confirm our baseline estimates using the traditional measure of bride price (reported in Tables 6 and 7 of the text). The estimates reported in column 1 indicate that females belonging to ethnic groups with a bride price ($I_e^{AltBridePrice} = 1$) positively and significantly benefit from the school construction program, while females from ethnic groups without a bride price ($I_e^{AltNoBridePrice} = 1$) do not. Column 2 reports the IV estimates. The estimated effects are almost exactly the same treatment effects as our main estimates in Table 6. Columns 3 and 4 report OLS and IV estimates using the continuous measure of bride price, $BP\,Continuous_e$. Although the estimates are slightly less precise, the same pattern is found. The effect of the school construction program on female education is greater for ethnic groups with a higher bride price measure.

Turning to Zambia, the patterns are similar. All the specifications confirm a larger effect of school construction on female education for bride price ethnic groups, and the estimates from column 6, where we use traditional bride price practices as an instrument, are almost identical to our baseline estimates reported in Table 7. The continuous measure indicates that females with higher bride price value-added benefit more from the school construction program in Zambia (columns 7 and 8). In these specifications, the estimated effect of the treatment on groups with the lowest bride price value (equal to zero) is actually negative and significant at the 10 or 5% level. This is potentially explained by the previous evidence that the districts to which schools in Zambia were allocated may have been negatively selected.

Testing for differential returns to education by ethnic group

An alternative explanation for the differential response of bride price ethnic groups to school construction is that the returns to education are higher. We test for this by checking whether the relationship between a women's level of education and either her employment status or her income is stronger for women from bride price cultures, relative to women from non-bride price cultures.

We estimate the following equation:

$$\begin{aligned} y_{iked} = {} & \alpha_d + \alpha_k + \beta_1 I_i^{Primary} + \beta_2 I_i^{Primary} \times I_e^{BridePrice} \\ & + \beta_3 I_i^{JuniorSecondary} + \beta_4 I_i^{JuniorSecondary} \times I_e^{BridePrice} \\ & + \beta_5 I_i^{SeniorSecondary} + \beta_6 I_i^{SeniorSecondary} \times I_e^{BridePrice} + \varepsilon_{kied} \end{aligned} \tag{A5}$$

where i indexes a woman between the ages of 25 and 45, e her ethnicity, d her district, and k her year-of-birth. y_{iked} is a proxy for income. We describe these in further detail below. $I_i^{Primary}$ is an indicator variable that equals one if individual i has completed primary school and attended junior secondary school, $I_i^{JuniorSecondary}$ is an indicator variable equal to 1 if an individual has completed junior secondary school and attended upper secondary school, and $I_i^{SeniorSecondary}$ is an indicator for having attended college. α_d denotes district fixed effects, which are intended to capture district-level differences in the labor market, and α_k denotes year-of-birth fixed effects. We cluster our standard errors at the ethnicity-level and report wild cluster bootstrapped p-values in the square brackets.

The estimates are reported in Appendix Table A17. For the Indonesian sample (columns 1–3), our outcomes are an indicator variable that equals one if the individual is employed, the natural log of the wage among those employed in the formal sector,[6] and the wealth index of a woman's household.[7] For the Zambian sample (columns 4–6), we also examine an employment indicator, as well as the wealth index that is calculated and reported directly in the Zambian DHS (now normalized to lie between 0 and 1), and an indicator variable if a woman's child (aged 0–5)[8] is categorized as being severely stunted.[9] Unfortunately, the Zambian DHS does not report wages.

[6] The Indonesia Intercensal data only report wages for formal sector employees.

[7] The wealth index is constructed by a principal components analysis including indicator variables for owning an automobile, television, radio, stove, buffet, bicycle boat, and motorboat. The wealth index is the first principal component of the factor analysis, and has then been normalized to lie between 0 and 1.

[8] Thus, the sample in column 6 is children aged 0–5.

[9] Severe stunting indicates that a child has a height more than three standard deviations below the mean of the

Overall, we find no evidence of a differential return to education for women belonging to bride price ethnic groups. Unsurprisingly, we find that more education is associated with a higher likelihood of employment, higher wages, and greater household wealth. However, we find no strong evidence that these relationships are systematically different for women from bride price ethnicities.[10] In Appendix Table A17, we also report the estimate of the predicted effect of the bride price indicator on the outcome of interest for a woman with the average value of education for the sample, which is a 63% chance of graduating primary school in Indonesia and a 32% chance of attending junior secondary school in Zambia. In all but one case (employment in Zambia), we cannot reject a zero effect of bride price for a woman with average education. In the case of employment in Zambia, the point estimate is negative, which is consistent with the fact that our model suggests that more educated bride price girls are negatively selected relative to more educated non-bride price girls.

Discrete choice model of schooling in Indonesia

In this subsection, we use a logit estimation procedure to asses the quantitative importance of the practice of bride price on female educational attainment in Indonesia. We focus on Indonesia rather than Zambia because the reliability of our estimates depends greatly on the quality of our measures of individuals' distances to schools. While we do not observe precise distances to schools in either country, we can approximate these distances using the number of schools in a district and the district's geographic area. This exercise is much less reliable in Zambia than in Indonesia, since districts in Zambia are much larger than those in Indonesia. The median district in Indonesia has an area of 1,857 km square, while the median district in Zambia is about five times as large at 9,203 km square. Additionally, in Indonesia, unlike Zambia, a secondary data source (the IFLS) provides us with an additional source of detailed distance data by documenting whether a more disaggregated geographic area – roughly a village – received a school in the INPRES expansion.

We take advantage of the fact that, according to our theoretical framework – see equation (1)

growth distribution for healthy children based on a study of six baseline countries chosen by the WHO – Brazil, India, Oman, Ghana, Norway and the USA.

[10]Although the coefficients of the interactions between educational attainment and bride price custom are not statistically significant, the point estimates suggest that returns to education may in fact be *lower* for girls in bride price groups. This pattern can be explained within our theoretical framework, where more lower-ability girls are educated in bride price groups than in non-bride price groups. While, in our model, ability only affects girls in the first period of their life, allowing lower ability girls to have lower returns to education later in life would account for these relationships in the data.

– a girl is educated if

$$a_i \geq a^*_{I_e}(k) \equiv \frac{k}{\gamma} - \frac{R^f}{1+r} - \frac{\Delta V_{i,e}}{1+r}\left(1 + I_e \frac{1-\gamma}{\gamma}\right),$$

which implies that the probability a girl completes primary school is

$$Pr(S=1|k, I_e) = Pr(a_i \geq a^*_{I_e}(k)) = 1 - G(a^*_{I_e}(k)).$$

Under the assumption that G is a type 1 extreme value distribution, we can estimate the parameters of $a^*(I_e, k))$ using a logit regression, where we approximate the cost of schooling k using the variation in the accessibility to schools induced by the INPRES school construction program.

Using the Intercensal data, we estimate:

$$Pr(S_i = 1) = 1 - G(\alpha_d + \alpha_k + \mathbf{X_{dk}\Gamma} + \beta DistanceDecline_d \times I_k^{Post} + \delta I_e^{BridePrice}) \qquad \text{(A6)}$$

where i indexes individuals, d districts, k birth cohorts, and e ethnicities. As before, $I_e^{BridePrice}$ is an indicator variable equal to one if ethnicity e traditional practice bride price, and I_k^{Post} is an indicator variable that equals one if those born in cohort k were young enough to benefit from the program (i.e., ages 2–6 in 1974 rather than 12–24). $DistanceDecline_d$ is a measure of the estimated decline in the populations' average distance to the nearest school in district d induced by the school construction program, whose construction we describe below. α_d are district of birth fixed effects, and α_k are birth cohort fixed effects. For the vector of controls, $\mathbf{X_{dk}}$, we include the same vector of controls as in Duflo (2001). Given that we are now examining the minimum distance to schools, we also add further controls that exactly mirror the baseline covariates in our original school construction specifications. These are birth-year fixed effects interacted with a child's minimum distance to school prior to the school construction program.

The ideal measure of $DistanceDecline_d$ would capture the decrease in distance to the nearest school due to the school construction program for each girl in our sample. However, this is not possible since we do not observe the exact locations of schools nor of individuals. Instead, we create a proxy measure of $DistanceDecline_d$ by assuming that schools and individuals are uniformly distributed in their district both before and after the program. We then calculate the average distance to the nearest school for individuals in each district using the pre-treatment number of schools and then using the post-treatment number of schools (which is the pre-treatment number of schools plus the schools that were built during the program). The variable $DistanceDecline_d$, which is the change in distance, is calculated as the average pre-treatment nearest distance minus the average post-treatment nearest distance. Thus, we expect the estimated coefficient for this variable to be positive: a greater reduction in distance is associated with more schooling.

Our rough measure of the decline in the distance to the nearest school is clearly imperfect since it is well-known that schools and populations are not evenly distributed within a district, and the extent to which the approximation is incorrect may vary systematically across districts and in a way that biases our estimates of interest. To test whether the results are robust to other ways of approximating distance to the closest school, we also use a second data source that, although available for a much smaller sample of individuals, has better information on actual distances to schools. This is rounds 3 and 4 of the Indonesia Family Life Survey (IFLS), which collects a school census at the community level (equivalent to an enumerator area in the Indonesia census), which is two levels more disaggregated than districts (i.e., one level smaller than a sub-district). The data also include information on whether a INPRES school was constructed in a community during the school construction program.

Using these data, we re-estimate our logit model above. Unlike in the previous specification, we now measure the school construction treatment at the community level and approximate the decrease in distance to the nearest school with a variable that equals the number of schools that were built in the community during the school construction program. The variable, which we denote $NewSchools_c$, takes on the values of 0, 1, 2 or 3. Our estimating equation is:

$$Pr(S_i = 1) = 1 - G(\alpha_c + \alpha_k + \mathbf{X_{ck}}\mathbf{\Gamma} + \beta NewSchools_c \times I_k^{Post} + \delta I_e^{BridePrice}), \qquad \text{(A7)}$$

where c indexes communities, and as before, k indexes cohorts, e indexes ethnic groups, α_c denote community fixed effects, α_k denote birth year fixed effects, I_k^{Post} indicates that a woman was young enough to be treated by the program (ages 2-6 in 1974 rather than 12-24), and $NewSchools_c$ is the number of newly-constructed schools built during the school construction program. The controls $\mathbf{X_{ck}}$ include the baseline Duflo (2001) covariates, as well as the number of schools in the village before the program interacted with cohort fixed effects. All covariates are allowed to vary by whether individuals belong to a bride price ethnic group.

In the equations above, δ can be interpreted as an estimate of $\frac{\Delta V_{i,e}}{1+r}\left(\frac{1-\gamma}{\gamma}\right)$. Thus, this term directly causes a difference between the educational attainment of girls from bride price and non-bride price ethnic groups. In addition, the two groups will also differ in the predicted effect of the school construction program, β. Thus, $\frac{\delta}{\beta}$ is the measure of the reduction in distance (equation (A6)) or the number of new schools (equation (A7)) that would have the equivalent effect to belonging to a bride price ethnicity on the gains that parents derive when their daughter completes primary school.

Appendix table A19 reports the results from estimating equations (A6) and (A7) using the Indonesia 1995 Intercensal data (columns 1-2) and the IFLS data (columns 3-4). Odd-numbered columns report the results without controls for additional ethnicity-level traits, while even-numbered columns include our set of additional ethnicity-level covariates. As the table shows, in both data sets, both the decline in distance and belonging to a bride price ethnic group positively affect the utility of a daughter completing primary school.

The estimates can be used to provide a sense of the quantitative importance of belonging to a bride price ethnic group for primary school completion relative to the decline in distance to the nearest school measure. This is done by calculating the effect of bride price on the utility from completing primary school in terms of the equivalent reduction in distance or increase in number of schools – i.e., by comparing δ to β and calculating δ/β. Based on this, belonging to a bride price ethnic group has an estimated effect that is equivalent to a reduction in the distance to the nearest school by 0.7–0.8 kilometers (from columns 1 and 2) or to 71%-85% of the effect of having an additional school constructed in a community containing 200–300 households (columns 3 and 4).

Appendix C: Appendix tables

Table A1: Ethnic groups in our Indonesian sample and whether bride price is practiced

Ethnicity	$I_e^{BridePrice}$	Ethnicity	$I_e^{BridePrice}$
BALINESE	0	ALORESE	1
CHAM	0	AMBONESE	1
DANI	0	BANGGAI	1
ENGGANO	0	BATAK	1
IBAN	0	BELU	1
JAVANESE	0	BUNGKU	1
KENYAH-KAYAN-KAJANG	0	DAWAN	1
KERAKI	0	GORONTALO	1
KUBU	0	ILI-MANDI	1
MARINDANI	0	KEI	1
MENTAWEIA	0	MACASSARE	1
MIMIKA	0	MALAYS	1
MINANGKAB	0	MANOBO	1
REJANG	0	MINAHASANS	1
SASAK	0	MUJU	1
SOROMADJA	0	MUNA	1
SUMBANESE	0	NIASANS	1
SUMBAWANE	0	PANTAR	1
SUNDANESE	0	ROTINESE	1
SUVANESE	0	SUGBUHANO	1
WAROPEN	0	TOBELORES	1
		TOMINI	1
		TORADJA	1

Notes: The table reports the ethnic groups in our Indonesian sample (1995 Indonesia Intercensal data), as well as the value of the bride price indicator variable for that ethnic group, where the indicator equals 1 if they practice bride price and 0 if they do not.

Table A2: Ethnic groups in our Zambian sample and whether bride price is practiced

Ethnicity	$I_e^{BridePrice}$	Ethnicity	$I_e^{BridePrice}$
BEMBA	0	BWILE	1
BISA	0	ILA	1
CHEWA	0	IWA	1
CHOKWE	0	LUNGU	1
KAONDE	0	MAMBWE	1
KUNDA	0	MBUNDA	1
LALA	0	PL.TONGA	1
LAMBA	0	SALA	1
LOZI	0	SHONA	1
LUANO	0	SOLI	1
LUCHAZI	0	TUMBUKA	1
LUNDA (LUA)	0		
LUNDA (LW)	0		
LUVALE	0		
NYANJA	0		
SHILA	0		
SWAKA	0		
TABWA	0		
USHI	0		

Notes: The table reports the ethnic groups in our Zambian sample (1996, 2001, 2007, and 2013 Zambia Demographic and Health Surveys (DHS)), as well as the value of the bride price indicator variable for that ethnic group, where the indicator equals 1 if they practice bride price and 0 if they do not.

Table A3: Correlations between bride price and other customs using a global sample

	(1)	(2)	(3)	(4)
	Dep Var: Bride Price Indicator			
Matrilineal	-0.118***			-0.153***
	(0.042)			(0.051)
Female Agriculture		0.109***		0.062
		(0.039)		(0.038)
Polygyny			0.335***	0.397***
			(0.034)	(0.041)
constant	0.534***	0.489***	0.233***	0.199***
	(0.015)	(0.023)	(0.031)	(0.036)
Number of observations	1,246	716	1,219	708

Notes: This table reports cross-ethnicity estimates of the relationship between the listed customs and the practice of bride price. The global data on traditional practices are from the *Ethnographic Atlas* (Murdock, 1967). Each column reports estimates from one regression where the dependent variable is an indicator variable that equals one if the ethnic group traditionally practices bride price. Robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Table A4: Contemporary bride price payments and traditional customs in Indonesia and Zambia

Panel A. Indonesian Sample		
IFLS (2000, 2007)	$I_e^{BridePrice} = 0$	$I_e^{BridePrice} = 1$
Number of observations	4,643	1,877
$\%(BP > 0)$	91%	92%
Mean (payment / per capita GDP)	37%	80%
Median (payment / per capita GDP)	4%	8.7%
Panel B. Zambian Sample		
ZFPS (2014)	$I_e^{BridePrice} = 0$	$I_e^{BridePrice} = 1$
Number of observations	524	179
$\%(BP > 0)$	77%	86%
Mean (payment / per capita GDP)	87%	205%
Median (payment / per capita GDP)	45%	57%

Notes: This table reports summary statistics on the prevalence of bride price practices and the size of bride prices paid in Indonesia and Zambia. The table draws on data from the 2001 and 2007 rounds of the Indonesia Family Life Survey and from the ZFPS (Fall 2014). Bride price payments greater than 100 times GDP per capita (4 observations in Zambia and 43 observations in Indonesia) are dropped as these extreme values are likely due to data entry error. Bride price data are from the *Ethnographic Atlas* (Murdock, 1967) and LeBar (1972) for Indonesia and from the *Ethnographic Atlas* (Murdock, 1967) and the *Ethnographic Survey of Africa* (Willis, 1966; Whiteley and Slaski, 1950; Schapera, 1938) for Zambia.

Table A5: Relationship between husband's education and wife's characteristics in Indonesia

	(1)	(2)
	Dependent Variable:	
	Indicator for Wife's Good Health	**Log Pre-Marital Assets of Wife**
Wife's Education:		
$I_i^{Primary}$	0.039***	0.183***
	(0.011)	(0.066)
$I_i^{JuniorSecondary}$	-0.010	0.342***
	(0.012)	(0.065)
$I_i^{SeniorSecondary}$	0.020	0.681***
	(0.015)	(0.068)
Husband's Education:		
$I_i^{H:Primary}$	0.006	0.178
	(0.018)	(0.228)
$I_i^{H:JuniorSecondary}$	0.021**	0.143**
	(0.011)	(0.070)
$I_i^{H:SeniorSecondary}$	0.002	0.267***
	(0.011)	(0.066)
Survey-Year FE, Birth-Year FE	Y	Y
F-stat for Husband's Educ. Coeff.	2.38	15.81
Number of observations	17,079	4,115
Adjusted R^2	0.049	0.267

Notes: This table reports the estimates of the relationship between the education of the husband and of the wife and the characteristics the wife. An observation is a married couple in the IFLS. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Table A6: First stage regression for instrumented effect of primary school completion on bride price values

	(1) **Dep Var: Indicator Variable for Completion of Primary School**
$Intensity_d \times I(Age_{1974} = 2)_i$	0.091
	(0.246)
$Intensity_d \times I(Age_{1974} = 3)_i$	0.028
	(0.183)
$Intensity_d \times I(Age_{1974} = 4)_i$	-0.392
	(0.266)
$Intensity_d \times I(Age_{1974} = 5)_i$	0.262
	(0.335)
$Intensity_d \times I(Age_{1974} = 6)_i$	0.183
	(0.648)
$Intensity_d \times I(Age_{1974} = 7)_i$	-0.256**
	(0.120)
$Intensity_d \times I(Age_{1974} = 8)_i$	-0.578
	(0.588)
$Intensity_d \times I(Age_{1974} = 9)_i$	-0.025
	(0.307)
$Intensity_d \times I(Age_{1974} = 10)_i$	-0.024
	(0.366)
$Intensity_d \times I(Age_{1974} = 11)_i$	0.206
	(0.231)
$Intensity_d \times I(Age_{1974} = 12)_i$	-0.376
	(0.427)
Baseline Covariates from Table 4, column 6	Y
Baseline Covariates from Table 6	Y
F-stat	3.04
Number of observations	264
Clusters	93
Adjusted R^2	0.259

Notes: This table reports the first stage estimates for the effect of primary school completion on the bride price payment received at marriage; the second stage estimate is reported in Table 4 of the main text. Following the specification used in Duflo (2001), we instrument for primary school completion with the interactions of treatment intensity and age in 1974 fixed effects. The sample is restricted to ethnic groups that practice bride price. Following Duflo (2001), we also restrict the sample to individuals born between 1950 and 1972 and allow the effect of school construction to vary by a child's age in 1974, restricting the effect to 0 if a child was older than 12 in 1974. d denotes a birth district. 'Baseline Covariates From Table 4' consist of ethnicity fixed effects, quadratics in marriage age and marriage year, and survey fixed effects. 'Baseline Covariates From Table 4' consist of district and cohort fixed effects, and the covariates included in Duflo (2001) (the interaction of cohort fixed effects with the number of school-aged children in the district in 1971, with the enrollment rate in 1971, and with the district-level implementation of a water and sanitation program under INPRES). Standard errors are clustered at the district level. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Table A7: Perceived determinants of bride price values in Zambia

Think about the factors that affect bride price today: what is the ... most important factor?					
		First	*Second*	*Third*	*Not listed*
Education	obs.	531	216	150	454
		39.3%	17.0%	13.0%	30.7%
Family values	obs.	222	268	206	655
		16.4%	21.1%	17.9%	44.6%
Good morals	obs.	197	278	210	666
		14.6%	21.9%	18.2%	45.3%
Virginity	obs.	131	179	177	864
		9.7%	14.1%	15.4%	60.9%
Tribe	obs.	96	115	185	955
		7.11%	9.1%	16.1%	67.8%
Age	obs.	39	93	139	1080
		2.9%	7.3%	12.1%	77.7%
Other	obs.	135	121	86	1009
		10.0%	9.5%	7.5%	73.0%

Notes: This table reports respondents' answers when asked "Think about the factors that affect bride price today. What is the [first, second, or third] most important factor?" An observation is a person, either male or female, who is a member of 715 couples from peri-urban Lusaka. The data were collected as a special module implemented in the first wave of the ZFPS in the Fall of 2014 (Ashraf et al., 2017).

Table A8: Perceived reasons for why bride price payments increase with education in Zambia

Reasons why bride price increases with education	(1) Unprompted reason	(2) Prompted reason	(3) Prompted not reason
Improves the bride's skills in the house	212	476	716
	15.1%	33.9%	51%
Improves the bride's knowledge and skills as a mother	171	600	625
	12.25%	42.98%	44.77%
Improves the woman's earning potential	207	713	468
	14.91%	51.37%	33.72%
Improves the literacy of children	85	725	573
	6.15%	52.42%	41.43%
Bride's parents should to be compensated for investments	777	448	179
	55.34%	31.91%	12.75%
Is associated with her parents being rich	104	507	785
	7.45%	36.32%	56.23%
Other reasons	25	-	645
	3.73%	-	96.27%

Notes: This table reports respondents' answers when asked why bride price increases with education. The data were collected by the authors as part of a special module implemented in the first wave of the ZFPS (Fall 2014). The sample comprises 715 couples interviewed in peri-urban Lusaka. An observation is an individual, either male or female. Respondents were first asked why bride price payments increase with education in Zambia unprompted and then were asked about specific reasons and could either agree or disagree. The unprompted responses are reported in column 1 and the prompted responses in columns 2 and 3.

Table A9: Balance statistics by bride price practice for the 1995 Indonesia Intercensus

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Bride Price		Non-Bride Price		Pooled Sample			
					Without FEs		With District & Treat. Cohort FEs	
	Mean	SD	Mean	SD	Difference	SE	Coefficient	SE
Primary school completion	0.639	0.480	0.605	0.489	0.034***	0.005	0.048*	0.026
$Intensity_d$	2.217	1.097	1.989	0.794	0.228***	0.009	-	-
I_k^{Post}	0.297	0.457	0.301	0.459	-0.004	0.005	-	-
Age	34.173	7.105	34.176	7.170	-0.003	0.077	-0.004	0.010
Muslim	0.601	0.490	0.900	0.300	-0.299***	0.004	-0.140***	0.050
Number school aged children in 1971	92,704	72,836	178,860	107,736	-86,156***	1,114	-	-
Water and sanitation program control	0.184	0.064	.0.158	0.048	0.026***	0.001	-	-
Matrilineal	0.090	0.287	0.109	0.311	-0.019***	0.003	-0.189***	0.050
Female Agriculture	0.038	0.191	0.035	0.185	0.003	0.002	-0.069**	0.034

Notes: This table reports balance statistics for the 1995 Indonesia Intercensal data. The sample has 84,648 observations–67,022 of which have known bride price status–and is used for the school construction analysis. It is composed of a treated group of individuals who were born between 1968 and 1972 and were therefore 2–6 at the time of school construction (1974) and an untreated group of individuals who were born between 1950 and 1962 and therefore were 12–24 at the time of the school construction. I_k^{Post} refers to the treated cohort, born from 1968–1972. $Intensity_d$ is the number of schools built in a district per 1,000 people in the school-aged population. Columns 1 and 2 present means and standard deviations for ethnicities that traditionally practice bride price. Columns 3 and 4 present summary statistics for non-bride price ethnicities. Column 5 reports the difference in means and column 6 reports the standard error of the difference. Column 7 reports the coefficient on bride price in a regression of the row-name variables on bride price practice, district of birth fixed effects, and treated or non-treated cohort fixed effects. Column 8 reports the standard error of the bride price coefficient, clustered at the district level. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Table A10: Effect of bride price by religion in Indonesia school construction sample

	(1) **Dep Var: Indicator Variable for Completion of Primary School**
$I_k^{Post} \times Intensity_d \times I(Muslim)_i \times I_e^{BridePrice}$	0.023
	(0.018)
$I_k^{Post} \times Intensity_d \times I(Muslim)_i$	-0.003
	(0.012)
$I_k^{Post} \times Intensity_d \times I(NonMuslim)_i \times I_e^{BridePrice}$	0.047***
	(0.015)
$I_k^{Post} \times Intensity_d \times I(NonMuslim)_i$	-0.011
	(0.010)
Baseline Covariates From Table 6, column 2	Y
Muslim Control and Interactions	Y
F-test (joint significance of bride price effects)	7.26
Number of observations	65,403
Clusters	240
Adjusted R^2	0.187

Notes: This table reports estimates of the effect of the INPRES school construction program on primary school completion, allowing for differences by religion (Muslim or not) and traditional bride price practice. Data on religion and educational attainment come from the 1995 Indonesia Intercensal Survey. Bride price data are from the *Ethnographic Atlas* (Murdock, 1967) and LeBar (1972). 'Baseline Covariates From Table 6' consist of district fixed effects, cohort fixed effects, and the controls from Duflo (2001) interacted with bride price status, as well as ethnicity by I_k^{Post} fixed effects, and ethnicity fixed effects interacted with $Intensity_d$. Duflo controls consist of cohort fixed effects interacted with the number of school-aged children in the district in 1971, cohort fixed effects interacted with the enrollment rate in 1971, and cohort fixed effects interacted with the regency-level implementation of a water and sanitation program under INPRES. These are the same controls as used in Duflo (2001). 'Muslim Control and Interactions' includes the indicator variable for whether an individual is Muslim, $I(Muslim)_i$, its interaction with cohort fixed effects, and its double interactions with $I_e^{BridePrice}$, $I_e^{NoBridePrice}$, and $Intensity_d$. The F-test tests the joint significance of $I_k^{Post} \times Intensity_d \times I(Muslim)_i \times I_e^{BridePrice}$ and $I_k^{Post} \times Intensity_d \times I(NonMuslim)_i \times I_e^{BridePrice}$. Standard errors are clustered at the birth district level. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Table A11: Bride price and the INPRES school expansion with effects by age in 1974

	(1) **Dep Var: Indicator Variable for Completion of Primary School**
$I_e^{BridePrice} \times I(Age_{1974} = 22 \text{ to } 24)_i$	0.004
	(0.014)
$I_e^{BridePrice} \times I(Age_{1974} = 19 \text{ to } 21)_i$	0.009
	(0.017)
$I_e^{BridePrice} \times I(Age_{1974} = 16 \text{ to } 18)_i$	0.013
	(0.011)
$I_e^{BridePrice} \times I(Age_{1974} = 10 \text{ to } 12)_i$	0.004
	(0.013)
$I_e^{BridePrice} \times I(Age_{1974} = 7 \text{ to } 9)_i$	0.007
	(0.010)
$I_e^{BridePrice} \times I(Age_{1974} = 4 \text{ to } 6)_i$	0.029***
	(0.010)
$I_e^{BridePrice} \times I(Age_{1974} = 1 \text{ to } 3)_i$	0.031*
	(0.016)
$I_e^{BridePrice} \times I(Age_{1974} = -2 \text{ to } 0)_i$	0.026**
	(0.011)
$I_e^{BridePrice} \times I(Age_{1974} = -5 \text{ to } -3)_i$	0.053***
	(0.020)
$I_e^{NoBridePrice} \times I(Age_{1974} = 22 \text{ to } 24)_i$	-0.004
	(0.011)
$I_e^{NoBridePrice} \times I(Age_{1974} = 19 \text{ to } 21)_i$	-0.011
	(0.010)
$I_e^{NoBridePrice} \times I(Age_{1974} = 16 \text{ to } 18)_i$	-0.011
	(0.010)
$I_e^{NoBridePrice} \times I(Age_{1974} = 10 \text{ to } 12)_i$	-0.010
	(0.010)
$I_e^{NoBridePrice} \times I(Age_{1974} = 7 \text{ to } 9)_i$	-0.008
	(0.010)
$I_e^{NoBridePrice} \times I(Age_{1974} = 4 \text{ to } 6)_i$	-0.015
	(0.011)
$I_e^{NoBridePrice} \times I(Age_{1974} = 1 \text{ to } 3)_i$	-0.002
	(0.012)
$I_e^{NoBridePrice} \times I(Age_{1974} = -2 \text{ to } 0)_i$	-0.004
	(0.012)
$I_e^{NoBridePrice} \times I(Age_{1974} = -5 \text{ to } -3)_i$	0.011
	(0.012)
Baseline Covariates	Y
F-test for Bride Price Coeff.	3.37
F-test for Non-Bride Price Coeff.	1.12
Number of observations	117,191
Clusters	253
Adjusted R^2	0.193

Notes: This table reports the coefficients and standard errors from the regressions plotted in Figure 3 (the estimated relationships between school construction and primary school completion for different 3-year birth cohorts). The 'Baseline Covriates' that are in the specification are the set of controls from column 2 of Table 6. Standard errors are clustered at the level of an individual's district of birth. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Table A12: Robustness checks for the effect of school construction on female education in Indonesia

	(1)	(2)	(3)	(4)	(5)	(6)
	Dep Var: Indicator Variable for Completion of Primary School					
	Census 2010	Census 2010	District Trends	District Trends	Placebo Regressions	Placebo Regressions
$I_k^{Post} \times Intensity_d$	0.011		0.009			
	(0.010)		(0.011)			
$I_k^{Post} \times Intensity_d \times I_e^{BridePrice}$		0.017**		0.018		
		(0.008)		(0.027)		
$I_k^{Post} \times Intensity_d \times I_e^{NoBridePrice}$		-0.015		-0.002		
		(0.011)		(0.016)		
$I_k^{PlaceboPost} \times Intensity_d$					-0.006	
					(0.005)	
$I_k^{PlaceboPost} \times Intensity_d \times I_e^{BridePrice}$						0.015
						(0.014)
$I_k^{PlaceboPost} \times Intensity_d \times I_e^{NoBridePrice}$						-0.004
						(0.007)
Baseline Covariates	Y	Y	Y	Y	Y	Y
Ethnicity FE	N	Y	N	Y	N	Y
Ethnicity FE $\times I_k^{Post}$	N	Y	N	Y	N	Y
Ethnicity FE $\times Intensity_d$	N	Y	N	Y	N	Y
District FE $\times I_e^{BridePrice}$	N	Y	N	Y	N	Y
Duflo Controls $\times I_e^{BridePrice}$	N	Y	N	Y	N	Y
Cohort FE $\times I_e^{BridePrice}$	N	Y	N	Y	N	Y
District-Specific Time Trends	N	N	Y	N	N	N
District by Bride Price Time Trends	N	N	N	Y	N	N
F-test (equality of coeff)		6.01		0.41		1.60
Number of observations	1,700,856	1,700,436	76,959	65,403	53,640	45,799
Clusters	263	263	255	240	247	232
Adjusted R^2	0.175	0.193	0.183	0.188	0.137	0.134

Notes: Columns 1 and 2 report estimates of the Indonesia school construction regressions using the Indonesia 2010 Census data. The education attainment data come from a ten percent sample of the 2010 Indonesia Census. Columns 3–4 re-estimate the specification in columns 1 and 2 of table 7, controlling for linear time trends, whose coefficients are allowed to vary either at the district level (column 3) or district by bride price level (column 4). Columns 5–6 report estimates of the Indonesia school construction regressions using a placebo treatment status instead of the true treatment status. The education attainment data are taken from the 2010 Indonesia census (columns 1 and 2) or the 1995 Indonesia Intercensal Survey (columns 3–6). Bride price data are from the *Ethnographic Atlas* (Murdock, 1967) and LeBar (1972). I_k^{Post} refers to the treated cohort, born from 1968–1972. The untreated cohort is born between 1950 and 1962. $Intensity_d$ is the number of schools built in a district per 1,000 people in the school-aged population. 'Baseline Covariates' are the same baseline controls as in Table 6. The 'Duflo controls' consist of cohort fixed effects interacted with the number of school-aged children in the district in 1971, cohort fixed effects interacted with the enrollment rate in 1971, and cohort fixed effects interacted with the regency-level implementation of a water and sanitation program under INPRES. These are the same controls as used in Duflo (2001). $I_k^{PlaceboPost}$ refers to the placebo treated cohort, who are aged 12–17 in 1974. The placebo untreated cohort is aged 17–24 in 1974. The subscript d indexes birth districts, k indexes cohorts, and e indexes ethnic groups. 'District-Specific Time Trends' denote district specific linear time trends. 'District by Bride Price Time Trends' denote district-specific linear time trends that are allowed to vary depending on whether an ethnic group practices bride price. Standard errors are clustered at the birth-district level. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Table A13: Robustness of estimates to the use of alternative ethnicity-level bride price measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Indonesia				Zambia			
	Dep Var: Indicator for Primary School Completion				Dep Var: Indicator for Primary School Enrollment			
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
$I_k^{Post} \times Intensity_d \times I_e^{AltBridePrice}$	0.016**	0.025**						
	(0.007)	(0.012)						
$I_k^{Post} \times Intensity_d \times I_e^{AltNoBridePrice}$	-0.006	-0.020						
	(0.033)	(0.014)						
$I_k^{Post} \times Intensity_d \times BP\,Continuous_e$			0.153*	0.260				
			(0.082)	(0.160)				
$I_k^{Post} \times Intensity_d$			-0.017	-0.046				
			(0.020)	(0.035)				
$Schools_{dt}/Area_d \times I_e^{AltBridePrice}$					0.011	0.043***		
					(0.011)	(0.014)		
$Schools_{dt}/Area_d \times I_e^{AltNoBridePrice}$					-0.012	-0.009		
					(0.018)	(0.011)		
$Schools_{dt}/Area_d \times BP\,Continuous_e$							1.189**	2.624**
							(0.538)	(1.144)
$Schools_{dt}/Area_d$							-0.062*	-0.130**
							(0.033)	(0.064)
Indonesia Covariates:								
Baseline Covariates from Table 6, column 2	Y	Y	Y	Y	N	N	N	N
Zambia Covariates:								
Baseline Covariates from Table 7, column 2	N	N	N	N	Y	Y	Y	Y
Number of observations	64,957	64,957	64,957	64,957	21,430	21,430	21,430	21,430
Clusters	240	240	240	240	71	71	71	71
Adjusted R^2	0.191	0.185	0.192	0.184	0.399	0.398	0.401	0.397
Mean of dependent variable	0.608	0.608	0.608	0.608	0.619	0.619	0.619	0.619
SD of dependent variable	0.488	0.488	0.488	0.488	0.486	0.486	0.486	0.486
F-test (equality of coeff)	0.43	12.84			1.12	18.21		

Notes: This table reports robustness tests of the differential effects of school construction using contemporaneous bride price payment data from Indonesia and Zambia to construct measures of an ethnic group's practice of bride price. Data on the value of bride price payments, which are from the IFLS and ZFPS, are used to estimate ethnicity-level bride price value fixed effects from a specification that accounts for educational attainment, a quadratic in marriage age, and a quadratic in marriage year. The ethnicity-level fixed effects are then matched to the ethnicities in the Indonesian Intercensus and the Zambia DHS and equation (7) and (10) of the paper are re-estimated using the alternative bride price measures. The variable $I_e^{AltBridePrice}$ is equal to one if an individual belongs to an ethnicity with a bride price fixed effect greater than or equal to the median value and zero otherwise. $I_e^{AltNoBridePrice}$ is equal to $1 - I_e^{AltBridePrice}$. $BP\,Continuous_e$ is the continuous measure of an ethnicity's bride price fixed effect. Columns 2, 4, 6, and 8 instrument for the contemporaneous measures, $I_e^{AltBridePrice}$, $I_e^{AltNoBridePrice}$, and $BP\,Continuous_e$, using the baseline bride price tradition indicator variable. The samples and control variables are the same as in Tables 6 (Indonesia) and 7 (Zambia) of the paper, except that, in the OLS specification, the definition of ethnic groups (and the resulting fixed effects) is slightly different due to the reporting of ethnic groups in the data used for the newly-constructed bride price measures; in addition, the controls that originally (in Tables 6 or 7) involve interactions with $I_e^{BridePrice}$ use the robustness bride price measure being used in that specification. In the IV specifications, which use the baseline bride price measures as instruments, we use the same ethnicity definitions and use the baseline measure for the controls that originally involve interactions with $I_e^{BridePrice}$. Standard errors, clustered at the level of the district of birth (in Indonesia) or the current district (Zambia), are reported in parentheses. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Table A14: Balance statistics by bride price practice for the pooled Zambia DHS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Bride Price		Non-Bride Price		Pooled Sample			
					Without FEs		With District & Cohort FEs	
	Mean	SD	Mean	SD	Difference	SE	Coefficient	SE
Primary School Enrollment	0.628	0.483	0.613	0.487	0.015**	0.007	0.007	0.012
Age	8.350	2.256	8.349	2.255	0.001	0.033	-	-
$Schools_{dt}/Area_d$	0.094	0.283	0.093	0.272	0.002	0.004	-0.002	0.002
Polygynous HH	0.245	0.430	0.122	0.327	0.123***	0.006	0.067***	0.012
Muslim HH Head	0.003	0.051	0.003	0.058	-0.001	0.001	0.000	0.001
Matrilineal	0.578	0.494	0.544	0.498	0.033***	0.007	-0.000	0.075
Female Agriculture	0.891	0.312	0.904	0.295	-0.013***	0.005	-0.033	0.027

Notes: This table reports balance statistics by bride price practice for the pooled 1996, 2001, 2007, and 2013 rounds of the Zambia DHS. The sample, which is that used in the school construction analysis for Zambia, includes 24,707 observations–22,402 of which have known bride price status–and comprises girls aged 5–12. Columns 1 and 2 present means and standard errors for the bride price group, and columns 3 and 4 report means and standard errors for the non-bride price group. Column 5 reports the difference in the means and column 6 reports the standard error of the difference. Column 7 reports the coefficient on bride price practice in a regression of the row-name variable on bride price practice and district and year fixed effects. Column 8 reports the standard error of this coefficient clustered at the district level. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Table A15: Robustness checks for the effect of school construction on female education in Zambia

	(1)	(2)	(3)	(4)
	Dep Var: Indicator Variable for Completion of Primary School			
	District Trends	District Trends	Placebo Regressions	Placebo Regressions
$Schools_{dt}/Area_d$	0.022		0.064	
	(0.126)		(0.074)	
$Schools_{dt}/Area_d \times I_e^{BridePrice}$		0.492***		0.294**
		(0.176)		(0.118)
$Schools_{dt}/Area_d \times I_e^{NoBridePrice}$		-0.127		-0.014
		(0.173)		(0.095)
$Schools_{d,t+1}/Area_d$			-0.063	
			(0.083)	
$Schools_{d,t+1}/Area_d \times I_e^{BridePrice}$				-0.227*
				(0.116)
$Schools_{d,t+1}/Area_d \times I_e^{NoBridePrice}$				-0.012
				(0.105)
Baseline Covariates	Y	Y	Y	Y
Age FE $\times I_e^{BridePrice}$	N	Y	N	Y
District-Specific Time Trends	Y	N	N	N
District by Bride Price Time Trends	N	Y	N	N
District FE $\times I_e^{BridePrice}$	N	Y	N	N
F-test (equality of coeff)		5.69		
Number of observations	22,191	22,191	12,370	12,370
Clusters	71	71	70	70
Adjusted R^2	0.400	0.402	0.393	0.393

Notes: Columns 1 and 2 re-estimate the effects of school construction in Zambia controlling for district-specific linear time trends, which are allowed to vary by bride price status. The regressions use girls aged 5–12 at the time of the survey in the pooled Zambia DHS surveys from 1996, 2001, 2007, and 2013. The subscript d indexes districts, k indexes cohorts, e indexes ethnic groups, and t indexes survey rounds. Columns 3-4 report estimates of the differential impact of present and future school construction in Zambia on bride price and non-bride price females. For these columns, we are unable to include the 2013 DHS, since we do not know how many schools will be built in the future. The treatment variable, $Schools_{dt}$, is the number of schools built in a district d by year t (the survey round of the DHS). This is normalized by the area of the district, $Area_d$. $Schools_{d,t+1}$ is the number of schools built by 2001 in 1996, the number of schools built by 2007 in 2001, and the number of schools built by 2013 in 2007. 'Baseline Covariates' are the same set of controls as in Table 7. 'District-Specific Time Trends' denote district specific linear time trends. 'District by Bride Price Time Trends' denote district-specific linear time trends that are allowed to vary depending on whether an ethnic group practices bride price. Standard errors are clustered at the district level. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Table A16: Bride price practice, wealth, sex ratio, and attitudes in Indonesia and Zambia

	(1) Wealth Index, 0-1	(2) Number of Children Born	(3) Fraction of Surviving Children that are Girls	(4) Fract Reasons Beating Wife Not Justified	(5) Fract Reasons Wife's Refusing Sex Justified
			Panel A. Indonesian Intercensus Sample		
$I_e^{BridePrice}$	-0.014	0.292*	-0.002		
	(0.016)	(0.078)	(0.005)		
	[0.488]	[0.054]	[0.650]		
Baseline Covariates	Y	Y	Y		
Unit of observation	HH	Wife	Wife		
Mean of dependent variable	0.391	3.020	0.489		
SD of dependent variable	0.252	1.930	0.338		
Number of observations	137,503	82,399	77,519		
Clusters	41	40	40		
Adjusted R^2	0.289	0.279	0.001		
			Panel B. Zambian DHS Sample		
$I_e^{BridePrice}$	0.005	-0.014	0.008	-0.025	-0.010
	(0.023)	(0.092)	(0.007)	(0.019)	(0.012)
	[0.848]	[0.894]	[0.268]	[0.294]	[0.614]
$I_i^{Female} \times I_e^{BridePrice}$				0.059	0.005
				(0.037)	(0.017)
				[0.212]	[0.838]
Baseline Covariates	Y	Y	Y	Y	Y
Unit of observation	HH	Wife	Wife	All	All
Mean of dependent variable	0.482	4.712	0.497	0.676	0.757
SD of dependent variable	0.341	2.530	0.286	0.379	0.313
Number of observations	17,403	15,893	15,216	21,607	21,600
Clusters	29	29	29	29	29
Adjusted R^2	0.354	0.377	-0.001	0.251	0.075

Notes: This table reports estimates of the relationship between various individual- or household-level characteristics and bride price status within Indonesia and Zambia. In column 1, the unit of observation is a household, in columns 2 and 3 it is women aged 25–45, and in columns 4–6, it is men and women aged 25–45. The dependent variable in column 1 is a wealth index. For Indonesia, the index is constructed using the first principle component using indicator variables for household assets. In Zambia, the DHS reports an assets variable. We normalize both variables to lie between zero and one. In column 2, the dependent variable is the total number of children born. In column 3, it is the number of surviving female children over the total number of surviving children born. In column 4, it is the fraction of reasons for which it was reported that a husband beating his wife is not justified. In column 5, it is the fraction of reasons for which it was reported that a wife's refusing to have sex was justified. For Indonesia, individual- and household-level data are from the 1995 Intercensal data. For Zambia, individual- and household-level data are from the 1996, 2001, 2007, and 2013 DHS. In Panel A, 'Baseline Covariates' include district fixed effects, the respondent's age and age squared (and the household head's age and age squared in the case of column 1), and an indicator variable equal to one if the household head (column 1) or respondent (columns 2 and 3) reports being Muslim. In Panel B, 'Baseline Covariates' consist of survey-year fixed effects, district fixed effects, the respondent's age and age squared (and the household head's age and age squared in the case of column 1), an indicator variable equal to one if the household head (column 1) or respondent (columns 2 and 3) reports being Muslim, and an indicator variable equal to one if the household is polygynous. Standard errors, clustered at the ethnicity level, are reported in parentheses. p-values obtained using the wild cluster bootstrap procedure appear in square brackets. *, **, and *** indicate significance at the 10, 5, and 1% levels according to the wild cluster bootstrap p-values.

Table A17: Testing for differential returns to education for bride price and non-bride price women in Indonesia and Zambia

	(1)	(2)	(3)	(4)	(5)	(6)
		Indonesia			**Zambia**	
	Employment	ln (Wage)	Wealth Index, 0-1	Employment	Wealth Index, 0-1	Child Stunting
$I_i^{Primary}$	-0.059***	0.280***	0.121***	0.046***	0.097***	-0.044***
	(0.004)	(0.014)	(0.004)	(0.013)	(0.006)	(0.007)
	[0.002]	[0.000]	[0.000]	[0.010]	[0.000]	[0.002]
$I_i^{JuniorSecondary}$	0.091***	0.827***	0.163***	0.024	0.234***	-0.050***
	(0.016)	(0.028)	(0.006)	(0.013)	(0.005)	(0.005)
	[0.000]	[0.000]	[0.000]	[0.108]	[0.000]	[0.002]
$I_i^{SeniorSecondary}$	0.177***	0.183***	0.053**	0.217***	0.180***	-0.076***
	(0.019)	(0.015)	(0.012)	(0.018)	(0.014)	(0.014)
	[0.000]	[0.000]	[0.022]	[0.000]	[0.000]	[0.002]
$I_i^{Primary} \times I_e^{BridePrice}$	-0.001	0.157	-0.005	0.007	-0.017	0.054
	(0.017)	(0.094)	(0.008)	(0.034)	(0.013)	(0.026)
	[0.952]	[0.330]	[0.600]	[0.856]	[0.298]	[0.140]
$I_i^{JuniorSecondary} \times I_e^{BridePrice}$	0.050	-0.152	-0.023*	0.023	-0.013	0.002
	(0.031)	(0.082)	(0.009)	(0.018)	(0.009)	(0.019)
	[0.348]	[0.264]	[0.062]	[0.270]	[0.222]	[0.994]
$I_i^{SeniorSecondary} \times I_e^{BridePrice}$	-0.026	-0.026	-0.025	0.018	-0.008	-0.015
	(0.054)	(0.042)	(0.024)	(0.026)	(0.026)	(0.030)
	[0.672]	[0.590]	[0.432]	[0.486]	[0.758]	[0.642]
$I_e^{BridePrice}$	-0.029	-0.020	-0.014	-0.051	0.020	-0.045*
	(0.027)	(0.032)	(0.108)	(0.041)	(0.019)	(0.021)
	[0.484]	[0.476]	[0.748]	[0.366]	[0.304]	[0.054]
Baseline Covariates	Y	Y	Y	Y	Y	Y
Effect of BP for women with average education for sample	-0.030	0.080	-0.018	-0.037***	-0.001	0.010
	(0.033)	(0.059)	(0.020)	(0.011)	(0.018)	(0.008)
F-test of bride price interact coeffs	4.96	1.39	3.30	1.10	1.35	2.33
Number of observations	88,049	12,547	88,039	19,671	17,145	21,946
Clusters	40	31	40	30	30	30
Adjusted R^2	0.106	0.506	0.413	0.099	0.498	0.053

Notes: This table reports estimates of the relationship between education and other outcomes for bride price and non-bride price women in Indonesia and Zambia. The samples consist of women aged 25–45, with the exception of column 6, which includes all children ages 0–5. Age controls consist of age and age squared. For Indonesia, educational attainment and labor market outcomes data are taken from the 1995 Intercensal data. For Zambia, educational attainment, labor market, and child health outcomes data are taken from the 1996, 2001, 2007, and 2013 Demographic and Health Surveys. Wage data in Indonesia are only reported for formal sector employees. $I_i^{Primary}$ is an indicator variable that equals one if individual i has completed primary school and attended junior secondary school, $I_i^{JuniorSecondary}$ is an indicator variable equal to 1 if an individual has completed junior secondary school and attended upper secondary school. In Zambia, $I_i^{SeniorSecondary}$ equals one if the woman has any post-secondary education, which is not necessarily at the university level, while in Indonesia $I_i^{SeniorSecondary}$ equals one if the person has any college or university education. 'Baseline Covariates' consist of controls for age and age squared and district by survey year fixed effects. The row 'Effect of BP for women with average education for sample' reports the estimated effect of bride price on the outcome of interest for a woman with the mean level of education for the sample. This is an 63.4% chance of primary school completion in Indonesia and a 31.7% chance of attending junior secondary school in Zambia. i indexes a woman between the ages of 25 and 45, e her ethnicity. p-values obtained using the wild cluster bootstrap procedure with 1000 draws are reported in square brackets. Standard errors, clustered at the ethnicity level, are reported in parentheses. *, **, and *** indicate significance at the 10, 5, and 1% levels according to the wild cluster bootstrapped p-values.

Table A18: Effects of school construction on male schooling

	(1)	(2)	(3)	(4)	(5)	(6)
	Indonesia			**Zambia**		
	Dep Var: Primary School Completion			Dep Var: Primary School Enrollment		
$I_k^{Post} \times Intensity_d \times I_e^{BridePrice}$	0.010	0.013	0.011			
	(0.010)	(0.010)	(0.010)			
	{0.007}	{0.003}	{0.003}			
$I_k^{Post} \times Intensity_d \times I_e^{NoBridePrice}$	0.020***	0.019**	0.017**			
	(0.007)	(0.008)	(0.008)			
	{0.004}	{0.009}	{0.009}			
$Schools_{dt}/Area_d \times I_e^{BridePrice}$				0.034*	0.015	0.018
				(0.017)	(0.022)	(0.055)
				{0.017}	{0.021}	{0.065}
$Schools_{dt}/Area_d \times I_e^{NoBridePrice}$				0.019	0.008	-0.010
				(0.014)	(0.017)	(0.058)
				{0.012}	{0.022}	{0.105}
Baseline Covariates	Y	Y	Y	Y	Y	Y
Polygynous Household	n/a	n/a	n/a	N	Y	Y
Muslim	N	Y	Y	N	Y	Y
Ethnicity Controls	N	N	Y	N	N	Y
F-test for difference in coefficients	0.62	0.29	0.31	0.56	0.11	1.24
Mean of dependent variable	0.724	0.724	0.724	0.595	0.586	0.586
SD of dependent variable	0.447	0.447	0.447	0.491	0.493	0.493
Number of observations	63,717	63,717	63,717	21,779	16,684	16,684
Clusters	247	247	247	71	71	71
Adjusted R^2	0.129	0.132	0.132	0.396	0.409	0.409

Notes: This table reports estimates of the specifications of columns 2, 5, and 6 of Tables 6 and 7 using a sample of males. The data for Indonesia are from the 1995 Intercensus. The data for Zambia are from the pooled 1996, 2001, 2007, and 2013 DHS. 'Baseline Covariates' consist of the control variables that are included in all specifications in the regressions of Tables 6 and 7. These are slightly different for the two samples and are described in detail in the notes of the two tables. 'Polygynous Household' is an indicator variable that equals one if the boy/man lives in a polygynous household. Since the Intercensus does not contain any households that report being polygynous, this control drops out of the Indonesia regressions. 'Muslim' is an indicator variable equal to one if the man self-reports being Muslim. 'Ethnicity Controls' consist of indicator variables for an individual belonging to an ethnic group that is matrilineal and that traditionally had female dominated agriculture. Standard errors clustered at the ethnicity level are reported in parentheses, while Conley standard errors with a distance cut-off of 400 km are reported in curly brackets. *, **, and *** indicate significance at the 10, 5, and 1% levels according to the clustered standard errors.

Table A19: Logit estimates of the determinants of female education in Indonesia

	(1)	(2)	(3)	(4)
	Intercensus Data		**IFLS**	
	Dep Var: Indicator for Primary School Completion			
$DistanceDecline_d \times I_k^{Post}$	0.821***	0.819***		
	(0.291)	(0.290)		
$NewSchools_c \times I_k^{Post}$			0.407***	0.400***
			(0.150)	(0.154)
$I_e^{BridePrice}$	0.648***	0.587***	0.288	0.338
	(0.196)	(0.163)	(0.309)	(0.302)
Baseline Covariates	Y	Y	Y	Y
Muslim Controls	N	Y	N	Y
Ethnicity Controls	N	Y	N	Y
Number of observations	65,319	65,319	5,825	5,823
Clusters	212	212	16	16
Adjusted R^2	0.152	0.154	0.293	0.293

Notes: This table reports estimates from logit model of the determinants of primary school completion for females in the 1995 Indonesia Intercensus. Following the empirical strategy of Duflo (2001), the sample consists of females born between either 1968 and 1972 or 1950 and 1962. I_k^{Post} refers to the treated cohort, born between 1968 and 1972, while the untreated cohort was born between 1950 and 1962. $DistanceDecline_d$ is the approximate change in the average minimum distance to a school in district d due to the school construction program. 'Baseline Covariates' include district fixed effects, cohort fixed effects, and the covariates included in Duflo (2001): the interaction of cohort fixed effects with the number of school-aged children in the district in 1971, with the enrollment rate in 1971 and with the district level implementation of a water and sanitation program under INPRES. Baseline covariates also include birth year interacted with the approximate average minimum distance to a school prior to the school construction program for the Intercensus data and the number of pre-treatment schools in the community interacted with birth year for the IFLS. 'Muslim Controls' consist of an indicator variable for being Muslim, its interactions with birth year and birth district fixed effects, and the triple interaction of the Muslim indicator, belonging to a treated cohort, and treatment intensity. 'Ethnicity Controls' consist of the interactions between ethnicity-level indicator variables for female dominated agriculture and matrilineality, being a treated cohort, and treatment intensity. The subscript d indexes birth districts or communities (in the IFLS), k indexes cohorts, and e indexes ethnic groups. Standard errors, clustered at the level of the district of birth, are reported in parentheses. *, **, and *** indicate significance at the 10, 5, and 1% levels.

Appendix D: Appendix figures

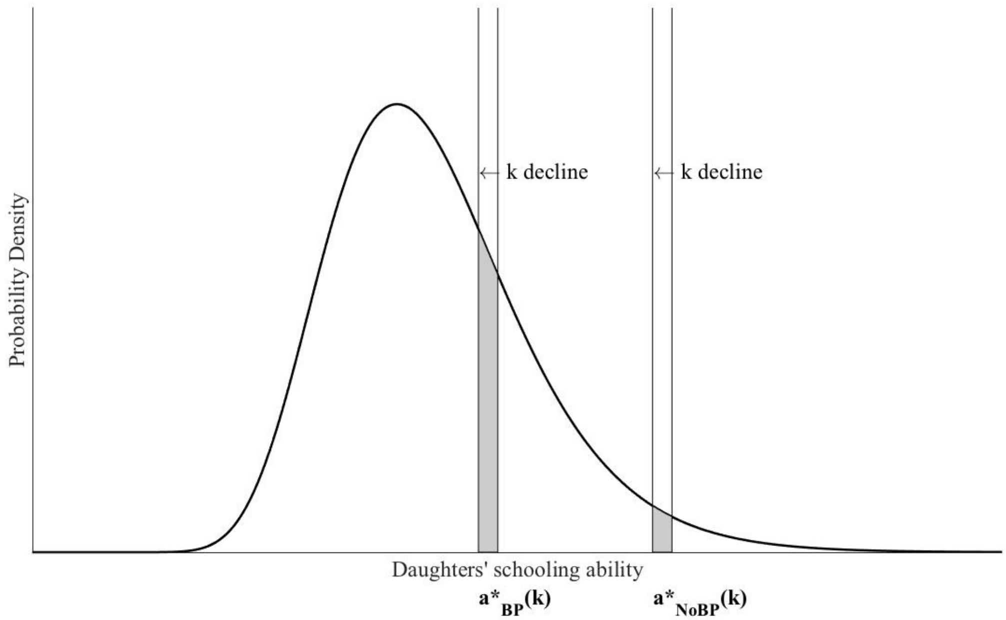

Figure A1: Example of dstribution of girls' abilities, schooling decisions, and declines in the cost of education in the model

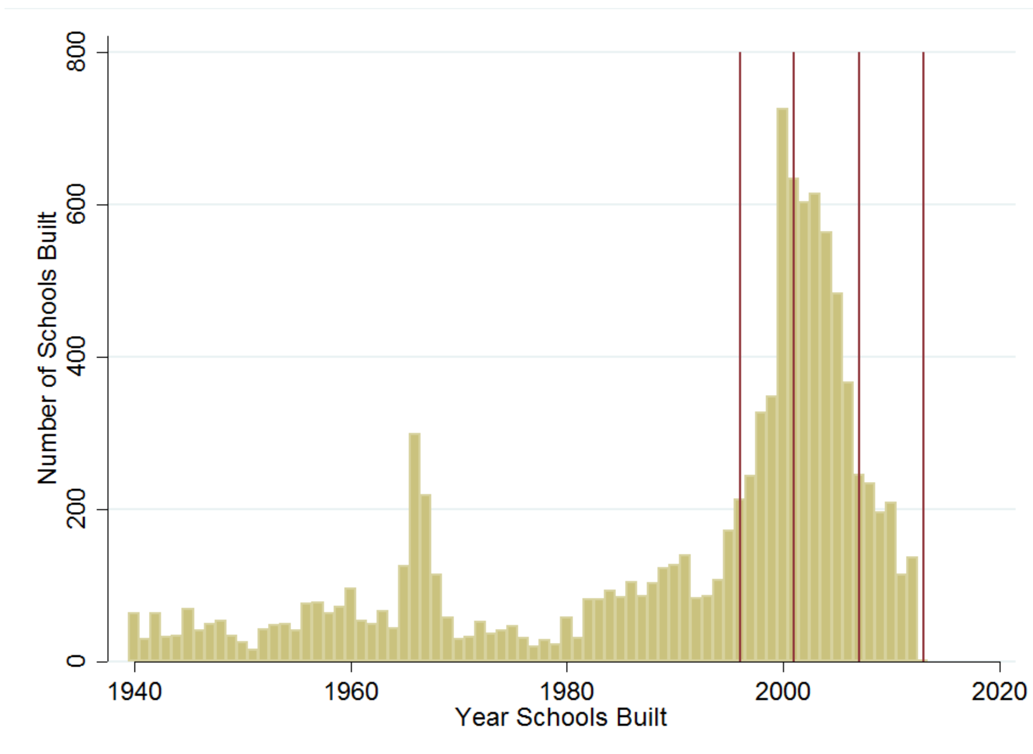

Figure A2: Number of schools constructed each year in Zambia (Ministry of Education, Government of Zambia). Vertical red lines indicate the years of the four rounds of the DHS.

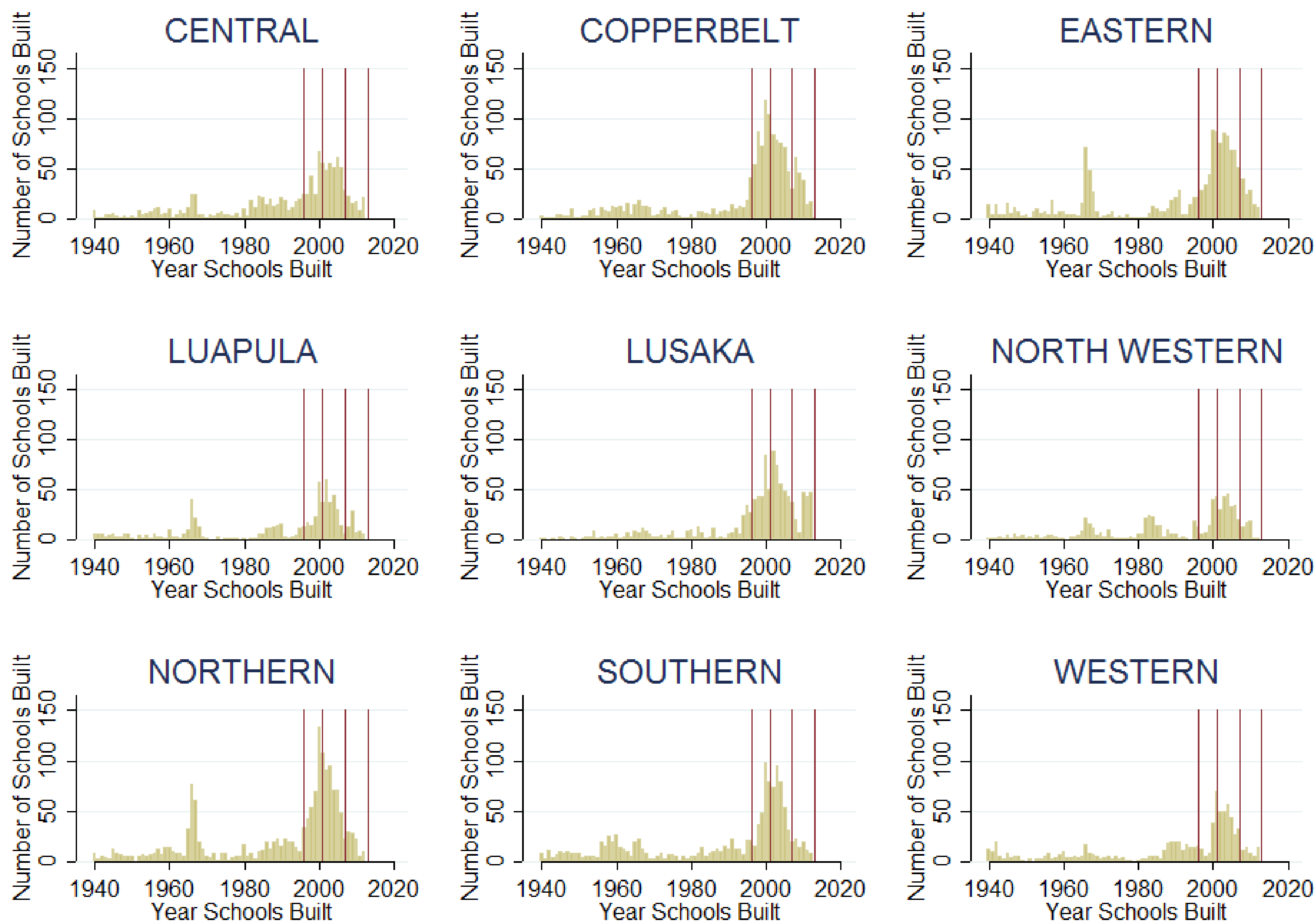

Figure A3: Number of schools constructed each year for each province in Zambia (Ministry of Education, Government of Zambia). Vertical red lines indicate the years of the four rounds of the DHS.

References

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics*, 2013, *128* (2), 155–194.

Ashraf, Nava, Erica Field, Giuditta Rusconi, Alessandra Voena, and Roberta Ziparo, "Traditional Beliefs and Learning About Maternal Risk in Zambia," *American Economic Review*, 2017, *107* (5), 511–515.

Chetty, Raj, John N Friedman, and Jonah E Rockoff, "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *American Economic Review*, 2014, *104* (9), 2593–2632.

Duflo, Esther, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 2001, *91* (4), 795–813.

Gordon, Raymond G., *Ethnologue: Languages of the World, 15th Edition*, SIL International, 2005.

LeBar, Frank M., *Ethnic Groups of Insular Southeast Asia. Volume 1: Indonesia, Adaman Islands, and Madagascar*, New Haven: Human Relations Area Files Press, 1972.

Minnesota Population Center, "Integrated Public Use Microdata Series, International: Version 6.1," 1995.

Murdock, George Peter, *Africa: Its Peoples and Their Cultural History*, New York: McGraw-Hill Book Company, 1959.

_ , *Ethnographic Atlas*, Pittsburgh: University of Pittsburgh Press, 1967.

Schapera, Isaac, *A Handbook of Tswana: Law and Custom*, Oxford: Oxford University Press, 1938.

Strauss, J., F. Witoelar, B. Sikoki, and A.M. Wattie, "The Fourth Wave of the Indonesian Family Life Survey (IFLS4): Overview and Field Report," 2009.

_ , **K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawati, and F. Witoelar**, "The Third Wave of the Indonesian Family Life Survey (IFLS): Overview and Field Report," 2004.

Whiteley, Wilfred and J. Slaski, *Bemba and Related Peoples of Northern Rhodesia*, London: International African Institute, 1950.

Willis, Roy, *The Fipa and Related Peoples of South-West Tanzania and North-East Zambia*, London: International African Institute, 1966.